DOCUMENT RESUME

ED 044 681 AL 002 656

AUTHOR Geoghegan, William H.
TITLE A Theory of Marking Rules.
INSTITUTION California Univ., Berkeley. Language and Behavior Research Lab.
SPONS AGENCY National Inst. of Mental Health (DHEW), Bethesda, Md.
REPORT NO WP-37
PUB DATE Oct 70
NOTE 38p.: Paper prepared for the Committee on Basic Research in Education Research Workshop on Cognitive Organization and Psychological Processes, Huntington Beach, Calif., August 16-21, 1970

EDRS PRICE EDRS Price MF-$0.25 HC-$2.00
DESCRIPTORS Cognitive Processes, *Communication (Thought Transfer), *Indonesian Languages, *Information Processing, Linguistic Theory, *Psycholinguistics, *Semantics
IDENTIFIERS Marking Rules, *Philippine Languages, Samal

ABSTRACT

In attempting to understand and describe the means by which human beings actually produce and interpret the message forms appropriate to a given domain of verbal behavior, it becomes apparent that the cognitive processes involved can reasonably be characterized as information processing phenomena. Throughout the production of a message form, an individual must select and encode the information that he wants to convey and take into account those additional informational items which, though not a part of the semantic content, are nevertheless necessary to its encoding. Similar considerations are necessary for the accurate interpretation (decoding) of messages received. The discussion presented here considers a particular type of natural information processing routine that appears especially significant in the semantic phases of message production and interpretation. Because of their close association with the phenomenon of semantic marking, routines of this type are here referred to as marking rules. A preliminary description of the structure and operation of such routines is followed by a detailed illustration drawn from a recent study of personal address terminology in a southern Philippine language (Samal). This example is then used as the basis for some further suggestions concerning the formal properties of marking rules in general. (Author/FWB)

A THEORY OF MARKING RULES*

William H. Geoghegan
University of California, Berkeley

Working Paper No. 37
Language-Behavior Research Laboratory
October 1970

The Language-Behavior Research Laboratory is supported by PHS Research Grant No. 1 R01 MH 18188-01 from the National Institute of Mental Health. This support is gratefully acknowledged.

*Paper prepared for the Committee on Basic Research in Education Research Workshop on Cognitive Organization and Psychological Processes, August 16-21, 1970, Huntington Beach, California.

A THEORY OF MARKING RULES[1]

William H. Geoghegan
University of California, Berkeley

Introduction

In attempting to understand and describe the means by which human beings actually produce and interpret the message forms[2] appropriate to a given domain of verbal behavior, it becomes apparent that the cognitive processes involved can reasonably be characterized as information processing phenomena. Throughout the production of a message form, an individual must select and encode the information that he wants to convey (the content of his communication) and take into account those additional informational items which, though not a part of the semantic content, are nevertheless necessary to its encoding. He must gather, organize and, in general, process such information in order to identify the message form consistent with his intent and capable of an appropriate interpretation. During the interpretive process as well, an individual must consider both the message form itself and any other information necessary for an accurate decoding, and process these items in order to generate an interpretation corresponding as closely as possible to the original semantic content.

In this discussion, I want to consider a particular type of natural information processing (IP) routine that appears to be especially significant in the semantic phases of message production and interpretation.

Because of their close association with the phenomenon of semantic marking (Greenberg 1966), I shall refer to routines of this type as marking rules.[3] Following a preliminary description of the structure and operation of such routines, I will provide a detailed illustration drawn from a recent study of personal address terminology in a southern Philippine language. This example will provide the basis for some further suggestions concerning the formal properties of marking rules in general. Since work in this area is still far from complete, my comments here should be regarded as highly tentative in nature and necessarily subject to a great deal of further verification.

The Nature of Marking Rules: Preliminary Considerations

Let me begin by describing a highly simplified version of an actual information processing routine of the type that can be represented by a marking rule. Assume, first of all, that a particular individual (call him Ego) has three types of personal names that he can use in addressing people that he knows. These name-types will be termed "pet name," "nickname," and "true name," symbolized PN, NN, and TN, respectively. For each person known to Ego, assume that there is at least one lexical item corresponding to each of these three types. His daughter's true name, for example, might be "Margaret," her nickname "Peggy," and her pet name "Punkin." For each potential addressee (or "Alter"), only one of the three possible forms will represent Ego's customary, normal, or expected mode of address. For the sake of simplicity, we can assume that this form is always Alter's nickname. Thus, for example, on most occasions when Ego addresses his daughter, he uses her nickname, "Peggy," rather than her pet name or true name. Such usage may be intended only to get her attention, to single her out within a larger group of people, to direct a remark in the course of a conversation, or to perform some other function of this general nature. Beyond this, it is a perfectly neutral form of address and does not directly communicate any information about Ego's momentary attitudes or feelings toward his daughter. His use of her nickname is highly predictable, in no sense unusual, and perfectly appropriate for situations in which no attitudinal information is to be conveyed. I shall refer to this type of usage as <u>unmarked</u>.

Now suppose that Ego wants to convey to his daughter (or to any other addressee) a feeling of affection for her, of intimacy, happiness, etc., through the use of an address form. In this situation, he shifts from his normal (or unmarked) usage and employs her pet name, "Punkin," as the appropriate form for communicating such information. In contrast, he uses his daughter's true name ("Margaret") to indicate anger, irritation with her behavior, or some other informational item of this general nature. Such items as "affection," "intimacy," "anger," and so forth, will be referred to as <u>marking cues</u> (or, more simply, <u>cues</u>); and forms such as <u>PN</u> and <u>TN</u> will be described as <u>marked</u> by one or more cues with respect to the normal, or unmarked form. Thus, for example, in Ego's address system <u>PN</u> (pet name) is marked by "affection" with respect to <u>NN</u> (nickname). A marked form can be said to <u>encode</u> the cues by which it is marked; while an unmarked form, on the other hand, encodes none of the cues potentially available.[4]

To describe the marking rule that represents an information processing routine of this sort, I will refer to items such as PN, NN, and TN as the outputs of the rule. Each output represents a potential result of applying the routine; the identity of the output finally chosen during any particular application will depend upon the identity of input information (an unmarked output and the cues to be encoded) taken to be in effect at that time. The final output will either be unmarked (in which case no cues have been encoded), or it will be marked by one or more of the available cues.

With regard to the process of applying the routine described above, a sequence of events of the following sort takes place. The marking rule is called into play whenever Ego has to address another individual with a personal name. There is a choice between three alternative outputs (name-types), and he must decide which one will appropriately encode the information that he wants to communicate. Ego first determines which output is unmarked for the current address situation. (NN is the only possible choice in this example.) If none of the cues that can be encoded by this rule represent part of the intent of his message, then the application will be terminated; and the final output will be the unmarked form, NN. But if he wants to communicate "affection," for example, then the output currently "in effect," so to speak, would change from NN to PN through the performance of an operation appropriate to the encoding of this item. At this point, no further encoding could take place, and PN would remain in effect as the final output for this application of the routine.

I shall use the term marking operator to refer to the information processing operation that is performed when a particular cue is to be encoded, and which produces the change in effective output that represents this encoding. Each of the cues specified for a particular marking rule will be associated with a single marking operator. A given operator, however, may be associated with more than one marking cue, so long as all such cues are encoded in precisely the same fashion through all possible applications of the rule.[5]

In more formal terms, a marking operator has the properties of a many-one function defined on the set of outputs specified for a given marking rule. Suppose we have two outputs $\underline{O}_i$ and $\underline{O}_j$ for some marking rule $\underline{M}$, and a marking operator $\underline{v}_k$ such that whenever $\underline{O}_i$ is in effect for some application of $\underline{M}$ and $\underline{v}_k$ is then applied (when one of its associated cues is encoded), the effective output changes to $\underline{O}_j$. Assume that $\underline{v}_k$ also has the capability of encoding its associated cues when the output $\underline{O}_u$ is in effect (producing a shift to $\underline{O}_v$), when $\underline{O}_x$ is in effect (producing a change to $\underline{O}_y$), and so forth. The set of potential applications of $\underline{v}_k$ can be represented as the mapping

$$\underline{v}_k(\underline{O}_i) = \underline{O}_j$$
$$\underline{v}_k(\underline{O}_u) = \underline{O}_v$$
$$\underline{v}_k(\underline{O}_x) = \underline{O}_y$$
etc.

With respect to the example we have been discussing, let $\underline{a}$ represent the marking operator associated with "affection," "intimacy," "happiness with Alter," etc., and let $\underline{x}$ represent the operator associated with "anger," "irritation," and so forth. The complete mapping performed by these two operators for the marking rule would then be:

$$a(NN) = PN$$

$$x(NN) = TN$$

A directed graph can be used as a convenient means of representing the mapping performed by each member of the set of marking operators upon the set of outputs specified for a given marking rule. Each output is represented by a single vertex of the graph, and each marking operator corresponds to one or more of the graph's arcs. If the operator v_k, for example, maps the output O_i onto O_j (i.e., is capable of causing such a change in effective output), then the graph for this rule would contain an arc (labeled by v_k) that is incident to both O_i and O_j and directed from the former to the latter. In Figure 1 such a graph is provided for the outputs and marking operators relevant to the example given earlier.[6]

$$\begin{tikzcd} PN \\ NN \arrow[u, "a"'] \arrow[d, "x"'] \\ TN \end{tikzcd}$$

Figure 1

Input Information

The input information for a marking rule in any given application includes those items that are entered directly from external sources during the performance of the routine and which are used to determine the identity of the final output for that application. (By "external sources," I am referring to memory, perception, other information processing routines, and so forth.) In addition to the marking cues that may be entered for encoding, one of the rule's <u>outputs</u> must also serve as an input item. Specifically, the encoding process itself cannot be initiated until it has been decided which of the rule's outputs is appropriately unmarked for the situation in which the rule is being applied. The identity of this output must necessarily be entered as the first item of input information in any given application. When there is only one output that can take this role (as in the example we have been discussing), the selection process is completely trivial; but there exist marking rules (one of which will be described shortly) in which any one of several different outputs has the potential of being unmarked, and where the process of choosing the one appropriate to a given situation is by no means trivial. In such cases the choice of an unmarked output can determine which combinations of cues are ultimately available for encoding, and which of the rule's other outputs actually encode them.

The selection of an unmarked output will normally require the application of one or more information processing operations external to the marking rule itself. These may involve nothing more complex than retrieving and entering the necessary information from memory; but such operations may also be organized into extremely complex IP routines which themselves require an extensive amount of input information for their performance.

Once the application of a marking rule has been initiated with the selection of an unmarked output, then the options open to the user at any subsequent point in the process (so long as additional encoding operations can take place) include not only the possibility of applying an available marking operator, but also the option of using no operator at all. Once an output O_i has been identified as in effect (it may or may not be unmarked), then it would normally be possible to terminate the rule's application and leave O_i in effect as the final output. In fact, what evidence there is concerning the operation of natural marking rules suggests that the probability of utilizing any given marking operator when it becomes applicable during the encoding process is very small indeed. Observations of Samal address behavior, for example, indicate that unmarked address forms (ones that were marked by none of the cues available in several obligatorily applied rules) characterize at least 90 per cent of everyday usage. The probability of applying any individual marking operator would consequently have to fall well below the figure of 0.10.

There are marking rules that specify obligatory cues -- ones that have to be encoded when they are identified and their associated operators are applicable -- but, even so, obligatory cues seem to be in effect for such rules in only a relatively small proportion of actual instances of use. In the Samal address system, one such cue comes into effect (and must be encoded) whenever the addressee is a hajji' (an individual who has made the pilgrimage to Mecca). The proportion of actual address situations in which this occurs is quite small over the long run, since well under one per cent of the Samal population hold this particular position. Another obligatory cue used in this system must be encoded whenever the originator of the address form (Ego) does not know the addressee's name (e.g., when the output corresponding to 'nickname' is in effect and Ego does not have a 'nickname' for Alter). Since most Samal address usage is directed toward people who are well known to an individual (family members, close friends, villagemates, and so forth), this cue is encoded in only a small proportion of the situations for which the address system is actually applied.

The low a priori liklihood of a marking cue being encoded when such an operation is possible does not stem from any structural characteristic of marking rules in general. Quite to the contrary, there is some reason to believe that the existence of this phenomenon is one of the necessary conditions for the development of IP routines of the marking rule type, and that other kinds of routines normally develop when this condition does not obtain. In other words, it appears that the structural and operational characteristics of marking rules are specifically adapted to the performance of information processing tasks for which individual items of input information have very low probabilities of occurrence.[7]

If marking operators are actually applied in only a small proportion of the situations for which they could be used, this will ultimately produce a highly skewed frequency distribution for the use of a marking rule's outputs. Considering a sufficiently large number of situations in which some output $\underline{O_i}$ is unmarked, for example, $\underline{O_i}$ will remain the final output far more often than those outputs $\underline{O_j}$, $\underline{O_k}$, ..., that are marked with respect to $\underline{O_i}$; and the frequencies with which $\underline{O_j}$, $\underline{O_k}$, ..., are actually used will generally decrease in rough proportion to their degree of marking with respect to $\underline{O_i}$ (the number of operators applied when they are marked with respect to this output). This ties in with my earlier statement that unmarked usage represents "normal," or "expected" behavior for a situation, and that marked forms are in some sense "unusual" or "unexpected." To a certain extent, it is also consistent with Greenberg's (1966) use of frequency distributions as evidence for semantic marking; and it certainly accords with my own more informal estimates of the frequencies of marked and unmarked usage in Samal address.

The Name-Type Rule in Samal Address

The simple example introduced earlier in this paper allowed us to discuss the principal entities (outputs, cues, operators, etc.) involved in IP routines of the marking rule type, as well as certain properties of the information used in their application; but there are additional characteristics of structure and operation that can be illustrated only through an example of much greater complexity than the one I have been treating up to now. I would also like to move away from hypothetical cases toward a consideration of more realistic instances of this phenomenon, in order to comment on the role of marking rules within more complex semantic systems. Accordingly, I will give a brief description of the address system used by one of my Samal informants[8] and proceed from there to a more detailed discussion of one of its constituent marking rules.

The address system employed by any individual Samal is a complex information processing routine used to produce and interpret address forms. I want to stress that an address form is not a particular message (an overt act). Rather, it is a message <u>form</u>: a conceptual representation of an infinitely extended class of potential messages, all of which share one or more basic features in common. The cognitive representation of these features constitutes the address form in question. Those features of immediate relevance to the address system include the lexemes that are used and their temporal ordering in a given message. (I will use square brackets to set off a given address form; the plus sign will indicate concatenation and the division between lexemes.) For example, the address form

[bapa' + hajji']

consists of two lexical items, bapa' and hajji', taken in this order. This form may be realized in actual speech as either bapa' hajji' or pa' hajji' (where pa' is a common contraction for bapa'), with a wide range of possible variation in stress, intonation contour, basic pitch level, vowel length, and so forth.[9] Similarly, the address form

[dakayu' danakan]

(literally, 'one sibling') consists of a single compound lexeme -- in this case a proname (or personal name substitute).

The address system described here is capable of generating over 200 different address forms, not considering the wide range of personal names available. These can be divided, however, into ten different address form-types (AFT's), according to the classes of lexical items represented in their constituents. There are four such classes, each of which contains lexical realizations for one of the following address elements:

A: 'address term' (bapa', babu', mbo', etc.)

T: 'honorific' (tuan and dayang)

G: 'positional title' or gallal term (hajji', 'imam, maharaja, etc.)

N: 'name', including personal names ('abdul, 'ali', hasan, etc.) and pronames (name substitutes) ('oto', nde', toto', etc.)

Each of the permissible AFT's contains either one, two, or three address elements in the order: A > T > G > N. As with address forms, I will use square brackets to delimit an AFT and the plus sign to indicate concatenation. Thus, for example,

[A + N]

denotes the address form-type that contains an 'address term' and a 'name', in this order. One of its possible lexical realizations would be the address form

[bapa' + 'abdul]

where bapa' is a realization for the address element A ('address term'), and 'abdul is a realization for N ('name'). The ten permissible AFT's are:

[A]

[A + T]

[A + T + G]

[A + G]

[A + N]

[T]

[T + G]

[G]

[G + N]

[N]

The production of an address form requires the application of two basic groups of operations. The first comprises what I will call the

AFT-selection routine. This procedure is employed to generate the AFT whose constituent elements will be given lexical realizations in the second stage of the production process. I should mention that any AFT is capable of carrying a certain amount of semantic content independent of its lexical realization. Specifically, the routine employed in selecting an AFT contains two marking rules that may be used to encode the degree and type of 'respect' ('addat) that Ego wants to convey to the addressee.

The second group of operations, referred to here as the lexical realization phase of production, consists of five separate information processing routines. There is one lexical realization routine for each of the address elements A, T, and G, and two such routines for the element N. Of the latter, one of the two is used when N appears in an AFT that also includes either A or G (there is no AFT containing both T and N); the other is applied when the AFT [N] has been chosen. In the former case, N can be realized only by a personal name (or by Ø, if a name is not known). When N is the only constituent of the AFT, however, it may be realized by either a personal name or by a proname. It is the IP routine used in the latter situation (the name-selection routine) and the marking rule used to perform one of its major operations that will be of particular concern to us in this discussion.

When an address form of type [N] is to be used, the selection of a realization for N proceeds in two phases. In the first of these, a marking rule (the name-type rule) is used to determine the type of personal name or proname that appropriately encodes the information (in addition to that carried by the AFT) that Ego wants to convey to the addressee. Once this decision has been made, a second operation must be performed in order to determine which lexical item is the proper realization for the name-type just selected.

There are seven types of personal names and pronames recognized by this informant (and other adult Samal speakers), each of which corresponds to an output of the name-type marking rule. These name-types are listed below, with brief descriptions and examples for each.

TN: 'ōn-na to'od 'his true name' (The personal name commonly recognized as an individual's actual, "given" name; e.g., 'abdulmuluk.)

TN': 'ōn-na to'od 'his true name'[10] (Proname-type corresponding to TN: e.g., lella ['man'/'male'], matto'a ['old person'], etc.)

NN: danglay-na 'his nickname' (A personal name frequently derived from an individual's 'true name' [e.g., 'abdulmuluk → muluk]; often the name by which he is referred to in the community at large.)

NN': danglay-na 'his nickname' (The proname-type corresponding to NN: e.g., dakayu' siali ['one younger-sibling'].)

PN: 'ugay'ugay-na 'his pet name' (Also, 'ugay'ugay-ku ma 'ia ['my pet name for him']. A personal name frequently derived from an individual's 'nickname' [e.g., muluk → llu', hakim → kki']. 'Pet names' are highly idiosyncratic, and may be derived from a number of sources other than 'nicknames': i.e., physical characteristics [e.g., sombeng ('harelip') → 'ombeng], past events, nonsense words, etc.)

PN': 'ugay'ugay-na 'his pet name' (The proname-type corresponding to PN: e.g., 'oto', nde', etc.)

T: This is a special proname-type that does not correspond directly to any type of personal name, but is rather derived from the Samal 'honorific'. It conveys a high degree of 'affect', and takes the realizations tuan (for males) and dayang (for females).

The personal name-types include TN, NN, and PN. Realizations for these are names of specific individuals and are determined in any particular address situation by the identity of the individual being addressed. The proname-types include three (TN', NN', and PN') whose realizations may serve under certain conditions as direct substitutes for personal names of the corresponding types (TN, NN, and PN, respectively). Lexical realizations for these are chosen by applying a code rule (Geoghegan, 1968, 1970) which requires consideration of the addressee's age group, sex, stage of development (for children), and relative age (for addressees of Ego's age group). Realizations for T (the fourth type of proname) are chosen according to the addressee's sex.

The marking rule used to select a name-type involves five different marking operators, four of which are optional in use and are associated with cues representing information concerned with certain attitudes and emotional states ('affect', 'anger', etc.). One operator (n) is obligatory when an associated cue characterizes the address situation. The five operators and their associated marking cues can be listed as follows:

a: 'positive affect':

'alasa ('aku) ma 'ia '(I) like him'/'(I) feel affectionate towards him'

kinōgan 'atay-ku ma 'ia 'my liver is made happy for him'

a': 'negative affect':

ngga'i ('aku) 'alasa ma 'ia '(I) do not like him'/'(I) do not feel affectionate towards him'

'ala'at 'atay-ku ma 'ia 'my liver is bad for him'

'akuddu' 'atay-ku ma 'ia 'my liver is upset/disturbed by him'

x: 'anger':

mag'amā 'aku ma 'ia 'I am angry at him'

nidugalan 'aku ma 'ia 'I am made upset/nauseated by him'

d: 'deference':

magmaltabat 'aku ma 'ia 'I defer to (show mild respect for) him' n: 'Alter (Alter's name) not known':

'insa' 'ia kinata'uhan ku 'he is not known to me'

ngga'i kata'uhan ku ma 'ia 'I do not know him.'

ngga'i kata'uhan ku X-na 'I do not know his X (name-type)'

(This list contains the descriptive phrases used by the informant to characterize the information that can be communicated through selection of a name-type.) I should point out that the cues associated with any one of these operators are essentially synonymous with one another, insofar as they connote particular attitudes or emotions; and they are, in fact, used interchangably by informants in statements regarding the use of personal address. (This may not be altogether clear from the rather literal translations of the Samal descriptive phrases.) For this reason, and to help simplify matters somewhat, I will refer to the cues associated with a given operator by a single collective gloss that stands for and roughly characterizes the information involved. Thus, for example, the gloss 'negative affect' will be used for the three cues associated with the marking operator a'. The mappings performed by these operators upon the outputs of the name-type rule are shown in the directed graph of Figure 2.

I stated earlier that a marking rule could have more than one unmarked output. This version of the name-type rule is a case in point, since it allows for either TN, NN, or PN to be used in this manner.[11] In any given application, the selection of an unmarked output is based on Ego's 'habitual' address usage to Alter (kabiaksahan pangōn, 'habitual means-of-naming'). If there has been a past history of interaction with the addressee sufficient for the establishment of a 'customary' or 'habitual'

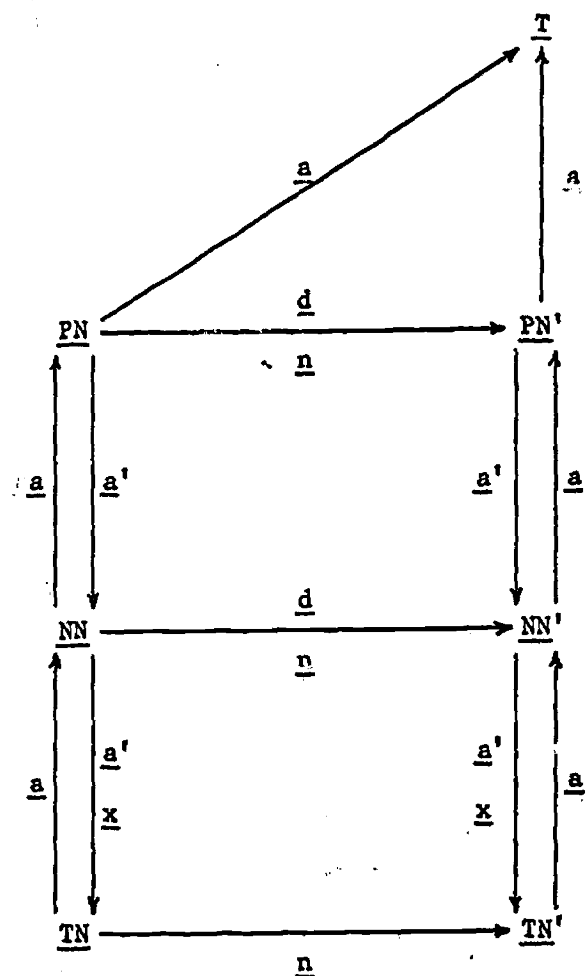

Figure 2

name, then the name-type corresponding to this form will be taken as unmarked. (Such an 'habitual name' is always interpreted as a realization of either TN, NN, or PN.) For addressees with whom past interaction has not been sufficient to allow for the growth of an 'habitual name', NN is taken as unmarked.[12]

Marking Sequences

The structure of the name-type rule allows more than one marking operator to be utilized in any given application. Suppose, for example, that Ego were to take NN as the unmarked output for a particular address situation. If he wanted to encode none of the cues available at this point in the process (those associated with the operators *a*, *a'*, *n*, *d*, and *x*), then NN would remain in effect as the final output, and application of the rule would cease. If he wanted to encode one of the 'positive affect' cues, on the other hand, use of the operator *a* would occasion a shift in effective output from NN to PN (see Figure 2). At this stage of the process, several additional encoding options would be available. Ego could choose to encode no further information (with PN becoming the rule's final output), or he could continue the application by encoding information associated with either *n*, *d*, or *a*.[13] (For reasons to be discussed shortly, *a'* would not be applicable once 'positive affect' had been encoded.) If Ego did not have a 'pet name' for Alter, then *n* would have to be applied (it is obligatory in such situations), producing a shift from PN to PN'. The same change in effective output would occur if he encoded 'deference' (*d*).[14] Once again, Ego would have the option of terminating his application of the rule, or of continuing with further encoding operations. The only available operator at this point in the process is a ('positive affect'), the use of which would produce a shift from PN' to T.

Another series of encoding operations (when NN is unmarked) might involve the initial use of n (if Alter's 'nickname' were unknown to Ego), causing a shift from NN to NN', followed by the encoding of 'anger' (x). The latter operation would produce a change in effective output from NN' to TN', at which point application of the marking rule would have to cease, leaving TN' in effect as the final output. (Although it appears that a is applicable at this point, 'positive affect' cannot be encoded simultaneously with 'anger'.) In general, a particular sequence of encoding operations (a series of successively applied marking operators) is possible only if we can discover in the graphical representation of the rule a path progression[15] corresponding to this sequence and originating with the effective unmarked output. In Figure 2, for example, we can find the path progression $\langle a, d, a\rangle$ originating at NN and terminating at T; and we also have the progression $\langle n, x\rangle$ originating at NN and terminating at TN'.

Given the set O of outputs for some marking rule M and a set V of marking operators v_i for M (where each such v_i is a many-one function defined on O), I will represent such a series of operations by a marking sequence V_k defined on O and V. In formal terms, a marking sequence V_k can be defined as an ordered n-tuple of marking operators v_i, where

$$V_k = \langle v_1, v_2, \ldots, v_{n-1}, v_n\rangle,$$

and for which the order of operators in V_k corresponds directly to the order of encoding operations in some potential application of $\underline{M}$. Although an ordered $\underline{n}$-tuple is normally taken to include at least two elements, we shall find it helpful to modify this practice in the representation of marking sequences. Specifically, we should allow for sequences that include only one operator (to represent applications for which only one marking operation occurs); and we shall also find it useful to allow for the limiting case of an "empty sequence" which contains no marking operators. The latter will be denoted

$$\underline{V}_0 = \phi$$

and will be used to characterize those applications of a marking rule in which no cues are encoded and for which the unmarked and final outputs are identical. Two of the marking sequences possible for the name-type rule when NN is unmarked are the ones described earlier: namely,

$\langle \underline{a}, \underline{d}, \underline{a} \rangle$ and

$\langle \underline{n}, \underline{x} \rangle$.

Permissible Applications

Given the set $\underline{O}$ of outputs and the set $\underline{V}$ of marking operators upon which a particular rule is constructed, it may be possible to define one or more marking sequences that correspond only to impermissible applications of the IP routine in question. An "impermissible application" is one that, while formally possible under the specification of $\underline{O}$ and $\underline{V}$, encodes a body of information that would be rejected by native speakers as "meaningless," self-contradictory, or in some other way semantically inappropriate. This will normally occur when the marking sequence representing a potential application contains the elements of a *cycle* (described below) or, in general, any pair of marking operators associated with semantically incompatible cues.

Referring to Figure 2, we can see that there are several potential marking sequences that correspond to path progressions beginning and terminating at the same output. The sequence <*a*, *a'*>, for example, leads from *NN* to *PN*, and then back to *NN*. (It also corresponds to a series of arcs beginning and ending at *TN*.) Similarly, the potential marking sequence <*x*, *a*> would, if *NN* were the unmarked output, simply map this output onto itself. I will refer to any marking sequence of this type as a *cycle*. When we find two or more marking operators that form a cycle, it will generally be the case that their associated cues are in some way mutually contradictory. That is, if a cue associated with one of these operators is in effect for a given situation, then cues associated with one or more of the other operators in the cycle cannot reasonably be in effect for that situation. All of the cycles that can be defined for the name-type rule contain either *a* and *a'*, or *a* and *x*. Some of the cues associated with *a* and *a'* are in direct "logical" conflict with one another (e.g., *'alasa ma 'ia* ['feel affectionate towards him'] and *nga'i 'alasa ma 'ia* ['do not feel affectionate towards him']), while others associated with these two operators conflict through their conceptual similarity (essentially synonymy) to "logically" incompatible cues. With regard to those associated with *a* and *x*, the conflict is of a more indirect nature. 'Anger' (*'amā*) and a sort of 'sick rage' (*dugal*), characterizing the cues associated with *x*, are normally taken to imply the concurrent existence of 'negative affect' (*a'*); which, in turn, implies the absence of 'positive affect' (*a*). As momentary attitudes or emotions, 'anger' and 'positive affect' are thus regarded as incompatible with one another.

A close inspection of Figure 2 should also indicate that even if the operators forming a cycle were utilized in the application of this rule, they would "neutralize" each other in terms of their effect on the identity of the final output. If *NN* were unmarked, for example, and if use of the operator *a* were followed by the use of *a*', then *NN* would be the final output of the rule -- that is, an output unmarked by cues associated with either of the two operators. Similarly, suppose that the marking sequence <*a*, *d*, *a*'> (which contains all elements of the cycle <*a*, *a*'>) were to be applied when *NN* was unmarked. The final output in this case would be *NN*', an output that is marked only by 'deference' (*d*) with respect to *NN* (assuming that *n* is not in effect). The cues associated with *a* and *a*' would be "cancelled out," so to speak, leaving 'deference' as the only informational item belonging to a reasonable interpretation for the use of *NN*' in this situation. In brief, even if Ego attempted to use this marking rule to encode cues associated with the elements of a cycle, these items of information could not be communicated by a single name-type because they would be "neutralized" during the encoding process itself.[16] (This phenomenon appears throughout those versions of the Samal address system that have been analyzed in detail.)

In addition to marking sequences of the type just described, there are others which, for one reason or another, represent impermissible applications of a given marking rule. Once again, such sequences generally contain two or more marking operators associated with contradictory or conflicting cues. Such conflicts are held to exist, for example, between 'deference' and 'negative affect', and between 'deference' and 'anger', for the Samal name-type rule. This follows from the fact that an overt display of 'deference' normally connotes a certain degree of 'respect' (*'addat*) on Ego's part, while the overt expression of 'negative affect' or 'anger' implies a definite *lack* of 'respect'. On several occasions, informants have stated explicitly that the encoding of such conflicting information is incorrect; and attempts to elicit these cues in the interpretation for such usage (e.g., the use of *NN'* when *PN* is the habitual or unmarked form) have consistently failed. (When *PN* is unmarked, the use of *NN'* is taken by informants to imply 'negative affect' and the fact that Ego has forgotten Alter's 'pet name' or 'nickname', an eventuality that would lead to the use of *n* rather than *d*.) Accordingly, it would seem appropriate to eliminate from the set of permissible marking sequences for a given rule all those which contain operators associated with such conflicting cues. In many respects, this restriction is identical to the one concerning sequences that contain the elements of a cycle (which also involve conflicting cues). The major difference is that cycles can be discovered on purely structural grounds (i.e., from the mappings performed by the various marking operators); while in the present case, structural criteria would normally be absent. (There is nothing about the mappings performed by *d* and *a'*, for example, that would lead us to conclude that their associated cues are incompatible with one another.)

While certain marking sequences should be ruled out as characterizing impermissible or inappropriate applications of a marking rule, there are other sequences of operators that should definitely be included as representing *necessarily* permissible applications. Referring once again to Figure 2, suppose we are given the marking sequence <*a*, *a*, *d*> as representing one appropriate use of the name-type rule when TN is the unmarked output. Assuming that the operator d is not made obligatory by the previous encoding of 'positive effect' (either once or twice), nor by the fact that TN is unmarked, then it should be possible to find applications of the rule for which it was permissible not to encode 'deference', and for which the marking sequence <a, a> represented the operations actually performed. Similarly, it should be possible to find applications for which it was permissible to encode 'positive affect' only once, and other applications for which no marking cues need be encoded and for which TN remained as the final output. In other words, given that <a, a, d> represents a permissible application when TN is unmarked, then each of the following marking sequences must also represent a permissible application under such conditions:

<a, a>

<a>

ϕ

(Recall that ϕ represents a "sequence" of operations in which none of the available marking operators is applied, and in which only an unmarked output is chosen.)

Phrasing the above in more general terms, if V_i is a marking sequence that represents a permissible application when O_u is unmarked, then each marking sequence constructed on V_i by taking the first m elements in order (where V_i contains n elements, and $n \geq m \geq 0$) also represents a permissible application of the rule when O_u is unmarked. Available evidence concerning the structure and operation of naturally occurring marking rules suggests that this should be taken as a characteristic of such information processing routines in general.

Table 1 contains a list of the thirty-five marking sequences that characterize permissible applications of this informant's version of the name-type rule. Each of these may be applied in conjunction with any of the three potentially unmarked outputs (PN, NN, and TN) for which the formal possibility exists. The sequence of operations represented by $\langle a, n \rangle$ (V_9), for example, may be applied in situations for which either TN or NN is unmarked. (In Figure 2, there is a corresponding path progression originating from each of these outputs.) This sequence may not be applied when PN is unmarked, because *a* maps PN onto *T*, and *T* does not fall within the domain of *n*. (That is, the operator *n* cannot be applied when *T* is the effective output. Figure 2 will show that there is no path progression corresponding to $\langle a, n \rangle$ originating at PN.)

The list shown in Table 1 was gathered through intensive elicitation sessions with the informant whose version of the name-type rule has been presented here. It should be noted that none of the marking sequences in this list are cycles, nor do any of them contain all the elements of a cycle. Similarly, there is no marking sequence in this set that contains both *d* ('deference') and *a'* ('negative affect') or both *d* and *x* ('anger'). It should also be noted that for every marking sequence V_i in this list, such that V_i contains n operators ($n \geq 1$), there is another sequence V_j in this set containing only the first m elements of V_i ($n \geq m \geq 0$) in the same order. With respect to the sequence V_{30}, for example, we have:

Table 1

$V_0 = \phi$

$V_1 = \langle a \rangle$

$V_2 = \langle a, a \rangle$

$V_3 = \langle a, a, a \rangle$

$V_4 = \langle a, d \rangle$

$V_5 = \langle a, a, d \rangle$

$V_6 = \langle a, a, d, a \rangle$

$V_7 = \langle a, d, a \rangle$

$V_8 = \langle a, d, a, a \rangle$

$V_9 = \langle a, n \rangle$

$V_{10} = \langle a, a, n \rangle$

$V_{11} = \langle a, a, n, a \rangle$

$V_{12} = \langle a, n, a \rangle$

$V_{13} = \langle a, n, a, a \rangle$

$V_{14} = \langle a' \rangle$

$V_{15} = \langle a', a' \rangle$

$V_{16} = \langle a', x \rangle$

$V_{17} = \langle a', x, n \rangle$

$V_{18} = \langle a', n, x \rangle$

$V_{19} = \langle a', n \rangle$

$V_{20} = \langle a', n, a' \rangle$

$V_{21} = \langle a', a', n \rangle$

$V_{22} = \langle x \rangle$

$V_{23} = \langle x, n \rangle$

$V_{24} = \langle n \rangle$

$V_{25} = \langle n, a \rangle$

$V_{26} = \langle n, a, a \rangle$

$V_{27} = \langle n, a, a, a \rangle$

$V_{28} = \langle n, a' \rangle$

$V_{29} = \langle n, a', a' \rangle$

$V_{30} = \langle n, a', x \rangle$

$V_{31} = \langle n, x \rangle$

$V_{32} = \langle d \rangle$

$V_{33} = \langle d, a \rangle$

$V_{34} = \langle d, a, a \rangle$

$\underline{V}_{30} = \langle \underline{n}, \underline{a}', \underline{x} \rangle$

$\underline{V}_{28} = \langle \underline{n}, \underline{a}' \rangle$

$\underline{V}_{24} = \langle \underline{n} \rangle$

$\underline{V}_{0} = \phi$

In other words, the set of permissible applications of this rule is consistent with the various constraints on such applications discussed in this section.

The Process of Interpretation

The existence of more than one potentially unmarked output for the name-type rule raises an important point concerning its use in both the production and interpretation of address forms. Specifically, neither the encoding nor the decoding process can take place without prior identification of the unmarked output in effect for a given application of the rule. This stems from the fact that the correspondence between an output and the information that it encodes varies according to the identity of the unmarked output chosen for a given application. Consider the various encodings shown in Table 2. If 'positive affect', for example, is the only item of information

Table 2

	Unmarked Output:			
Cues Encoded:	TN:	NN:	PN:	
'positive affect' ($\langle \underline{a} \rangle$)	NN	PN	T	
no cues encoded (ϕ)	TN	NN	PN	Final Output
'negative affect' ($\langle \underline{a}' \rangle$)	*	TN	NN	

* The operator $\underline{a}'$ cannot be applied when TN is in effect.

to be communicated, then it is appropriately encoded by a 'nickname' when TN is unmarked, by a 'pet name' when NN is unmarked, and by a proname of type T when PN is the unmarked output. Without prior identification of the unmarked form, no decision could be made regarding the proper method of encoding this item.

With respect to the interpretive process, suppose that Ego addressed Alter with the latter's 'nickname'. If Alter were unaware of the output that Ego originally took as unmarked, then he would have no way of determining which of three conflicting interpretations was the correct one. More precisely, a 'nickname' is marked by 'positive affect' when TN is unmarked; it is marked by 'negative affect' when PN is unmarked; and it encodes no cues at all when NN itself is the unmarked output. Not only do the three possible interpretations differ, but two of them ('positive affect' and 'negative affect') are in complete conflict with one another. If Alter could decide which output Ego took as unmarked, then he would be able to provide the message with a more or less unambiguous interpretation; but should he disagree with Ego in this identification, then misinterpretation would be the inevitable result.

In other words, for this marking rule to be at all effective in communicating information, there must be some procedure for identifying the unmarked output each time the rule is applied, during both production and interpretation; and this procedure must normally lead to agreement between Ego and Alter. During the production process, as I mentioned earlier, a rather brief series of operations is employed for this purpose. Ego would first attempt to identify his 'habitual name' (kabiaksahan pangõn) for Alter.

If such a name existed, then it would be entered as an item of input information to the address system, and would be analyzed to determine its corresponding name-type. The result would then be entered as the unmarked output for this application. If there were no 'habitual name', on the other hand, then NN would be taken as unmarked.

This series of operations is also applied, however, during the initial phases of the _interpretive_ process; even though it is basically a _production_ routine. When employed in this manner, the routine is applied as it is assumed to have been used during production of the address form: that is, from Ego's point of view. Alter would first attempt to identify Ego's 'habitual name' for him; and, given that this can be done, determine the corresponding name-type to be entered as the unmarked output. If such a name could not be identified, then Alter would take NN as unmarked. Following the performance of these operations, the decoding of cues could take place. Suppose Alter had been addressed by a 'pet name', and that Ego's customary address usage to Alter was the latter's 'nickname'. In the initial phase of interpretation (what I shall refer to from now on as the _production phase_), Alter would identify NN as the output Ego probably took as unmarked. He would also analyze the address form actually used and note that it contained a realization for PN (taken to be the final output). Comparison of these two outputs within the marking rule would indicate that 'positive affect' had been encoded; and this would be the interpretation assigned to Ego's original message. Without identification of the unmarked output, the address form would have had an ambiguous interpretation, since a name of type PN could encode either no cues (when PN is unmarked), 'positive affect' (when NN is unmarked), or "extreme" 'positive affect' (when TN is unmarked and the operator a is applied twice in succession).

While the production phase of the interpretive process involves a relatively simple IP routine for the name-type rule, in other segments of the Samal address system the production phase can become quite complex. Selection of an address form-type (AFT), for example, requires that one of two "AFT marking rules" be applied. (These rules are used to encode information concerning the degree and type of 'respect' ['addat] that Ego wants to communicate to Alter.) Determining the unmarked output for either one can involve the use of two additional IP rules (a code rule and a marking rule) as well as a number of subsidiary operations for the input and analysis of necessary information. Items concerning Ego's 'habitual' address form for Alter, Alter's age group relative to Ego's, Alter's status as a hajji' (one who has made the pilgrimage to Mecca), and so forth, may all become relevant to the choice of an unmarked AFT.

Although items such as these ('habitual' forms, relative age, and so forth) are employed in the production process and affect the identity of the final address form, they cannot realistically be described as representing any portion of the primary intent of such a message. That is, they are not items that Ego would normally encode in an address form simply for the purpose of communicating them to Alter. Alter can reasonably be expected to know whether or not Ego has an 'habitual' address form and name for him, the actual identity of such forms, his age relative to Ego's, his own status as a hajji', and (with regard to other portions of the address system) his sex, age group, kin relation to Ego, and so forth. Rather, such information has the primary role of establishing a framework or context in which other, more immediately salient items of information can be encoded or decoded. Consider the nature of information used in the name-type rule. It includes such items as 'affection' for Alter, 'happiness' with him, attitudes of 'deference', 'dislike', 'anger', and so forth. These are concerned with relatively private emotional and attitudinal states; and they are normally encoded in an address form only when Ego explicitly wants such information to have an effect on Alter's current state of knowledge, on his attitudes, and on the behavior that might be expected to follow from such. When actually encoded in an address form, items of this type exemplify what I would prefer to call the semantic content of the message -- information that Ego is deliberately attempting to convey to a particular individual.

I should emphasize, however, that by far the largest proportion of actual address usage in Samal is completely unmarked, and devoid of semantic content in the present sense of the term. (Most instances of personal address usage have what appears to be a metacommunicative function: they signal the opening of communication, direct messages to specific individuals, serve to emphasize portions of a complex utterance, and so forth.) For the relatively small proportion of cases in which content information is actually encoded in an address form, this is invariably done through the use of marking rules, a fact that holds for every version of the address system elicited during the course of this study.[17]

It is particularly important to maintain a strong distinction between content and context information (where the latter includes items such as an 'habitual name', relative age, etc.) if we are to understand adequately this type of communicative process. These two types of information differ not only in regard to communicative intent, but they play entirely different roles during the production and interpretation of address forms. This is best illustrated by the interpretive process, where context items are used during the production phase (to "generate" unmarked outputs for the various marking rules) in much the same manner as during the actual production of an address form, and where content items represent information retrieved from a message form during the decoding phase of interpretation. The fact that context information is normally shared between the parties in an address situation is from this standpoint not accidental, but a necessary precondition to effective communication. Given that content information in the address system is encoded and decoded through the use of marking rules, and that Ego and Alter must agree on the identity of the unmarked output in effect for a given rule during any address situation, it follows that there must also be agreement between the two individuals on the identity of those context items that are used in selecting the unmarked output. If such items were not shared, then disagreement on the identity of the unmarked output would be likely to follow; and this, in turn, could easily lead to miscommunication (where the information encoded by Ego does not correspond to that decoded by Alter).

In summary, an adequate understanding of the Samal address system -- as a device for interpersonal communication -- cannot be obtained unless we make a careful distinction between those items of information that may be used as part of a message's content and those which are used to establish a context for the encoding and decoding of content items. It is important to note, also, that the bulk of the information normally used in the production of an address form is of the context variety, is necessarily shared between the two parties to a communication in normal address usage, and does not represent a portion of the sender's communicative intent. As I suggested earlier, this implies that most instances of personal address usage involve forms that are relatively free of semantic content (that is, which are used to transmit little or no internally encoded information). When content information is encoded during an application of the address system, however, this is invariably done through the use of one or more marking rules.

FOOTNOTES

1. The data upon which this paper is based were collected during a study of personal address terminology among the Balangingi' Samal (a Muslim group of the southern Philippines). Fieldwork in the Philippines was supported by the National Institute of Mental Health (research grant number MH-13089-1). During subsequent periods of research and analysis, support was generously provided by the Institute of International Studies and the Committee on Research (both of U.C., Berkeley), the Office of Education (Department of HEW: grant number OEG-9-9-140281-0038(057), John Gumperz principal investigator), and by an NIMH grant (USPHS MH-18188-01: Paul Kay principal investigator) for the Language-Behavior Research Laboratory at Berkeley. The assistance of these agencies is gratefully acknowledged. Brent Berlin, Roy D'Andrade, Charles Frake, Paul Kay, and Robert Randall have provided many helpful comments and criticisms during informal discussions; and while this paper has benefitted greatly from their assistance, its errors and omissions are solely my own.

2. For purposes of this discussion, a message form will be regarded as a cognitively localized configuration of information that may be realized by, or represent, any one of a number of alternative, though equivalent messages. The message itself is an overt act that may vary on a number of attributes not directly relevant to the identity of the message <u>form</u> which it realizes (e.g., in terms of certain paralinguistic features).

3. A preliminary, and highly informal discussion of this topic was presented in an earlier paper (Geoghegan 1969). For a more general treatment of information processing systems and rules, and a tentative formalization for a theory of marking rules, see Geoghegan (1970).

4. The actual use of an unmarked form does not imply that the attitudes or feelings represented by available cues do not currently exist, but only that Ego has not chosen to communicate such information through his choice of a name-type. Moreover, if Ego had decided to communicate "anger" by addressing Alter with the latter's true name, this would not necessarily imply that the attitude was true in some objective sense, but rather that Ego had simply chosen, for whatever reason, to communicate such information to the addressee. (In reprimanding his daughter, for example, Ego might address her as "Margaret" <u>[TN]</u> in order to communicate "anger," even though actually amused by her misbehavior.) What we are concerned with here is the process by which an individual goes about encoding information once he has decided to communicate it to another individual -- not with the truth or falsity of what he wants to say.

5. In naturally occurring IP routines of the marking rule type, it generally appears to be the case that cues which are encoded in the same manner are conceptually, or semantically, quite similar to one another. In an example presented later in this paper, there are several cues that in native usage correspond to minor variations of a more basic concept that might be labeled 'positive affect'. Each of these cues is encoded in precisely the same way, and they are regarded by informants as essentially synonymous in personal address.

6. I should mention that such a graph does not normally contain all the information necessary for the complete specification of a given marking rule. It describes only the mapping performed by each of the rule's operators, but it does not indicate the range of permissible applications of the rule itself.

7. A detailed discussion of this point would take us far beyond the limited scope of this paper. For a more complete treatment of the acquisition and development of information processing rules and a discussion of possible cognitive mechanisms relevant to this process, see Geoghegan (1970).

8. I conducted a study of the semantics of personal address among the Balangingi' Samal in 1966-67 in the barrio of Tagtabon (Tictauan Island), approximately six miles east of Zamboanga City center (Mindanao). Nine informants were interviewed at length on this subject, and the address systems of six were selected for detailed analysis. The present discussion concerns the system used by a woman who was 56 years old at the time of the study. While there is extensive variation between informants as to the details of personal address, the basic structure of the system is the same in all cases; and each of the informants utilized a marking rule, similar in most respects to the one given here, for the selection of name-types.

9. These paralinguistic features can be semantically important, but their selection depends on the use of IP routines that are relevant to verbal behavior in general and which lie outside the address system proper.

10. The name-types TN', NN', and PN' are sometimes referred to by an expression such as:

X 'ia 'ōn-na to'od bang ngga'i kata'uhan ku 'ōn-na (to'ód).

"X is his 'true name' when I do not know his (true) name."

11. Other versions of the rule (elicited from different informants) may vary on this point. Name-type rules used by children, adolescents, and young adults usually allow for only one unmarked output (NN). Rules used by older adults permit two or three outputs to be unmarked. (The address system, and especially the name-type rule, continues to develop in the direction of greater complexity until an individual is about 50 or 60 years old.)

12. The name by which an individual is normally known and referred to in a community is usually described as his 'nickname' (danglay). Although Ego's 'habitual' usage to a given addressee may be a 'pet name' or 'true name', this form would not generally be used in reference, except to persons who customarily address the individual in this manner.

13. In using the name-type rule, it is possible to encode either 'positive affect' (a) or 'negative affect' (a') more than once in a single application. This has the effect of communicating a more intense or stronger version of the attitude in question. In situations for which TN is unmarked, for example, 'positive affect' could be encoded once (with NN as the final output if no other cues were relevant), twice (with PN as the final output),

or three times (with _T_ as the final output). The use of _T_ in an address form would communicate a higher "degree" of 'positive affect' than would _PN_; and _PN_ would indicate a more intense attitude than _NN_.

14. 'Positive affect' and 'deference' represent a fairly frequent combination of marking cues in Samal usage. The act of communicating these items (through address or otherwise) is referred to by the term _bijjak_ ("to cajole") and is often performed by an individual when he desires the recipient of his message to do him a large favor.

15. A path progression is composed of a sequence of arcs, each consecutive pair of which contains arcs adjacent to a single vertex (output) such that one arc is incident into the vertex and the other is incident out of the vertex. (See Busacker and Saaty 1965:27.)

16. This is not meant to imply that the Samal are incapable of irony in personal address (i.e., the encoding of contradictory attitudinal or emotional cues), but rather that irony cannot be effectively communicated through the use of a single marking rule. One can convey 'positive affect' and 'negative affect' simultaneously, for example, by applying the operator _a_ in the name-type rule, and by applying an operator encoding 'negative affect' in one of the marking rules used to generate paralinguistic features of an utterance.

17. I suspect that this might be true for address systems in general. It certainly holds for the Samal system, and seems to be the case for the American English and Bisayan systems as well. There is no _a priori_ reason why marking rules should be the _only_ type of IP routine capable of encoding content items. This phenomenon seems rather to stem from the relatively infrequent use of marked address forms (quite likely universal to such systems), and the fact that IP routines of the marking rule type are particularly well adapted to the efficient encoding of infrequently used items of information. (See Geoghegan 1970.)

REFERENCES CITED

Busacker, Robert G., and Thomas L. Saaty

1965 Finite graphs and networks: an introduction with applications. New York, McGraw-Hill.

Geoghegan, William H.

1968 Information processing systems in culture. In Paul Kay, ed., Exploraticns in mathematical anthropology. Cambridge, MIT Press. (In press.)

1969 The use of marking rules in semantic systems. Working Paper No. 26. Language-Behavior Research Laboratory. University of California, Berkeley.

1970 Natural information processing rules: formal theory and applications to ethnography. Unpublished Ph.D. dissertation. Anthropology Department, Stanford University.

Greenberg, Joseph H.

1966 Language universals. The Hague, Mouton & Co.